

SEQUENCE LISTING

(1) GENERAL INFORMATION

- (i) APPLICANT: Tang, Y. Tom  
Corley, Neil C.  
Guegler, Karl J.
- (ii) TITLE OF THE INVENTION: HUMAN SMN-LIKE PROTEIN
- (iii) NUMBER OF SEQUENCES: 4
- (iv) CORRESPONDENCE ADDRESS:
  - (A) ADDRESSEE: Incyte Pharmaceuticals, Inc.
  - (B) STREET: 3174 Porter Dr.
  - (C) CITY: Palo Alto
  - (D) STATE: CA
  - (E) COUNTRY: USA
  - (F) ZIP: 94304
- (v) COMPUTER READABLE FORM:
  - (A) MEDIUM TYPE: Diskette
  - (B) COMPUTER: IBM Compatible
  - (C) OPERATING SYSTEM: DOS
  - (D) SOFTWARE: FastSEQ for Windows Version 2.0
- (vi) CURRENT APPLICATION DATA:
  - (A) APPLICATION NUMBER: To Be Assigned
  - (B) FILING DATE:
  - (C) CLASSIFICATION:
- (vii) PRIOR APPLICATION DATA:
  - (A) APPLICATION NUMBER:
  - (B) FILING DATE:
- (viii) ATTORNEY/AGENT INFORMATION:
  - (A) NAME: Billings, Lucy J.
  - (B) REGISTRATION NUMBER: 36,749
  - (C) REFERENCE/DOCKET NUMBER: PF-0482 US
- (ix) TELECOMMUNICATION INFORMATION:
  - (A) TELEPHONE: 650-855-0555
  - (B) TELEFAX: 650-845-4166
  - (C) TELEX:

(2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 238 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (vii) IMMEDIATE SOURCE:
  - (A) LIBRARY: BRSTNOT24
  - (B) CLONE: 3769729

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```

Met Ser Glu Asp Leu Ala Lys Gln Leu Ala Ser Tyr Lys Ala Gln Leu
 1           5           10           15
Gln Gln Val Glu Ala Ala Leu Ser Gly Asn Gly Glu Asn Glu Asp Leu
 20           25           30
Leu Lys Leu Lys Lys Asp Leu Gln Glu Val Ile Glu Leu Thr Lys Asp
 35           40           45
Leu Leu Ser Thr Gln Pro Ser Glu Thr Leu Ala Ser Ser Asp Ser Phe
 50           55           60
Ala Ser Thr Gln Pro Thr His Ser Trp Lys Val Gly Asp Lys Cys Met
 65           70           75           80
Ala Val Trp Ser Glu Asp Gly Gln Cys Tyr Glu Ala Glu Ile Glu Glu
 85           90           95
Ile Asp Glu Glu Asn Gly Thr Ala Ala Ile Thr Phe Ala Gly Tyr Gly
100          105          110
Asn Ala Glu Val Thr Pro Leu Leu Asn Leu Lys Pro Val Glu Glu Gly
115          120          125
Arg Lys Ala Lys Glu Asp Ser Gly Asn Lys Pro Met Ser Lys Lys Glu
130          135          140
Met Ile Ala Gln Gln Arg Glu Tyr Lys Lys Lys Lys Ala Leu Lys Lys
145          150          155          160
Ala Gln Arg Ile Lys Glu Leu Glu Gln Glu Arg Glu Asp Gln Lys Val
165          170          175
Lys Trp Gln Gln Phe Asn Asn Arg Ala Tyr Ser Lys Asn Lys Lys Gly
180          185          190
Gln Val Lys Arg Ser Ile Phe Ala Ser Pro Glu Ser Val Thr Gly Lys
195          200          205
Val Gly Val Gly Thr Cys Gly Ile Ala Asp Lys Pro Met Thr Gln Tyr
210          215          220
Gln Asp Thr Ser Lys Tyr Asn Val Arg His Leu Met Pro Gln
225          230          235

```

## (2) INFORMATION FOR SEQ ID NO:2:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2426 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (vii) IMMEDIATE SOURCE:

- (A) LIBRARY: BRSTNOT24
- (B) CLONE: 3769729

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```

TCTTTCATAG AGACTAAAGT TATTCAGCAG GCAGCAAAAT AATCTACTTA AGTCCTGCCT      60
TTCTTTTTTC ACTTAAAAAA GTGGGTGTGA TAATATCCAG GCTAGCTAGC TGACTAGCTC      120
CCCGGGCAGT CTATGATAAT CAGAGATAGT CAATTTATTA GGCTGTTTTG CTGAATAAGC      180
TGGTTCATAA GGAGGCAGGG GTCAAGTCAC TTGTCTCATA TATTACAGTG GCTCTCTGCA      240
TCCCCGAAAC GCCTTCCTTC AGTAAGCAGA GTGCTTGAGT GCACCCCAT T GACCTGCTG      300
ATATGTAGAT CACAACNCCT GATGCTTCCT GGAATTGCCG ATTACTGTAA CTGCTGCCCA      360
TCTGTGCGATG AAGGAGCAGT TTCAGAACTC AGACTTGAGG GAGGAAAAGT AATTAATGGT      420
GCCCCGCGTT CCTCCCTTCC CCCTCGCCGC CGACCGAGTT CTTCTTTTTC AGACCGGGTC      480
GCCTTGCTGT CGTCGCGGTG ATTTTCCTGC TACTGCTACT GCTGCTGCTG CCACCGCCAC      540
TACCACTGGG CTCATTTGCC CCGACCCCTT CCCGCCGCC CGCCCCCAGC CCCACACAAG      600
ATGTCAGAGG ATTTAGCAAA GCAGCTGGCA AGCTACAAAG CTCAGCTCCA GCAAGTTGAA      660
GCTGCATTAT CTGGAAATGG AGAAAATGAA GATTTGCTAA AATTGAAGAA AGATTTACAA      720

```

```

GAAGTTATAG AACTAACCAA AGACCTTCTG TCAACTCAAC CTTCTGAGAC GCTTGCAAGT 780
TCAGACAGTT TTGCTTCTAC TCAACCTACT CATTTCATGGA AAGTAGGAGA CAAGTGTATG 840
GCAGTCTGGA GTGAAGATGG ACAGTGTTAT GAAGCGGAGA TTGAGGAGAT AGATGAAGAA 900
AATGGCACCG CTGCAATCAC CTTTGCTGGT TATGGCAATG CTGAAGTGAC TCCACTGTTG 960
AACCTCAAGC CTGTAGAAGA AGGAAGGAAG GCAAAGGAGG ACAGTGGCAA CAAACCCATG 1020
TCAAAAAAAG AAATGATTGC CCAGCAGCGT GAATATAAAA AGAAGAAAGC TTTGAAAAAA 1080
GCTCAGAGAA TAAAAGAACT TGAGCAGGAA AGAGAGGACC AGAAAGTGAA ATGGCAACAA 1140
TTCAACAACA GAGCCTATTG TAAAAACAAA AAAGGCCAGG TAAAGAGGAG TATTTTTGCT 1200
TCACCTGAGA GTGTGACTGG TAAAGTTGGA GTAGGAACCT GTGGAATTGC TGATAAACCT 1260
ATGACACAAT ATCAAGATAC CTCTAAATAC AATGTCAGGC ATTTGATGCC TCAATAATCA 1320
GAAAAAAGTGT TGGATTTTCAT CTCTGCAGGG CTTTACATTT ACCTTTTATAT CTTTATATTT 1380
TTCTAAAGGT AAATTATTTG TTAGATGAGT AAGCAAGATA CCATTGTCGT CATTGGTTGG 1440
CTTCAGTAGA ATGAAACGTG AAGAAATTGC ATTTGATAAC TGCTATTTCAT TTAACTTTTC 1500
TCATTATCAG TACCACGGTT CCCTCAAAGT TTGTTGAATA AAGCAACTTT TGTAGATGCT 1560
GTTTCATACA GCACTTAGAT GAATTATTGA TCTTCCTAAT ATCAGGCGCC TACTTAACCT 1620
ATGGTGTGTA CTTTTTGTA GTTGTAACCT GAAATTTTCA GATGCTTTGA ACTTGACACA 1680
TACTCTAGCA ATTCATTGGA ACACCAAGGC AAAAACACCA ACCTGCTAAA AGAGATCTTT 1740
TCATTTTTCT TATTTTCAGC TTTAAACTT AGCTGTCGTT CAGTTAAGCT TAAAGATAGG 1800
TTAATTTGTA AATGGCAAAG TTTGTTTTGA GGTTTTTCCT CAATAACTTG TTTCCCAGGC 1860
CTATTAGGCC ATCTCTAAAA TTGATCTAGC TGTTTTATTT TTATGTACTC TTAGTTTTAT 1920
GTAAGAAACC TTAGGATGAG CTCCCTTTTC TAAGGTGTTT TTGTTTTTTG TATGTTTGCT 1980
TTTTTCCTGT TTTTGTGTTT TTCCATTTAC GGCAGTGGTA CCATGTTTTG GATGTGTGAT 2040
GTTTATATGG GAGAACAAAA AGCTGATGTA TAGCCCTGTA TACAGTGTAG ATACTATTTT 2100
TGTAATAAACA CAAGGCTAAA TTAATGAACA AGAATACTGA ATATTTTCATC ATTAAAAATT 2160
TCTTGATATT CTTGTGCATT AATCTGACGA TAATTTCCCT GTATATTATG TTCATTTAGC 2220
TGTTTGTAAT TTTTGTTAAT TAGATCAGGT TGCTGTCATT TGTGTTGTA AGTGAACATC 2280
ATCACAGTTA TCCTGAGTTG AGTTTAAGCC AAATACATGC ATAGAAAAGG GTCTTCCTAT 2340
TAATGGAAGA AGGTAATTTT TAGGATGTGT ATTATTTTCAG TTTTGTATGT TTAACTTTTA 2400
TTAAATAAAG TGTTTTTAAA ATCTCC 2426

```

## (2) INFORMATION FOR SEQ ID NO:3:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 288 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (vii) IMMEDIATE SOURCE:

- (A) LIBRARY: GenBank
- (B) CLONE: 1857114

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

```

Met Ala Met Gly Ser Gly Gly Ala Gly Ser Glu Gln Glu Asp Thr Val
1           5           10           15
Leu Phe Arg Arg Gly Thr Gly Gln Ser Asp Asp Ser Asp Ile Trp Asp
20           25           30
Asp Thr Ala Leu Ile Lys Ala Tyr Asp Lys Ala Val Ala Ser Phe Lys
35           40           45
His Ala Leu Lys Asn Gly Asp Ile Cys Glu Thr Pro Asp Lys Pro Lys
50           55           60
Gly Thr Ala Arg Arg Lys Pro Ala Lys Lys Asn Lys Ser Gln Lys Lys
65           70           75           80
Asn Ala Thr Thr Pro Leu Lys Gln Trp Lys Val Gly Asp Lys Cys Ser
85           90           95
Ala Val Trp Ser Glu Asp Gly Cys Ile Tyr Pro Ala Thr Ile Thr Ser
100          105          110
Ile Asp Phe Lys Arg Glu Thr Cys Val Val Val Tyr Thr Gly Tyr Gly

```

```

      115              120              125
Asn Arg Glu Glu Gln Asn Leu Ser Asp Leu Leu Ser Pro Thr Cys Glu
      130              135              140
Val Ala Asn Ser Thr Glu Gln Asn Thr Gln Glu Asn Glu Ser Gln Val
      145              150              155              160
Ser Thr Asp Asp Ser Glu His Ser Ser Arg Ser Leu Arg Ser Lys Ala
      165              170              175
His Ser Lys Ser Lys Ala Ala Pro Trp Thr Ser Phe Leu Pro Pro Pro
      180              185              190
Pro Pro Met Pro Gly Ser Gly Leu Gly Pro Gly Lys Pro Gly Leu Lys
      195              200              205
Phe Asn Gly Pro Pro Pro Pro Pro Pro Leu Pro Pro Pro Phe Leu
      210              215              220
Pro Cys Trp Met Pro Pro Phe Pro Ser Gly Pro Pro Ile Ile Pro Pro
      225              230              235              240
Pro Pro Pro Ile Ser Pro Asp Cys Leu Asp Asp Thr Asp Ala Leu Gly
      245              250              255
Ser Met Leu Ile Ser Trp Tyr Met Ser Gly Tyr His Thr Gly Tyr Tyr
      260              265              270
Met Gly Phe Arg Gln Asn Lys Lys Glu Gly Lys Cys Ser His Thr Asn
      275              280              285

```

## (2) INFORMATION FOR SEQ ID NO:4:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 294 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (vii) IMMEDIATE SOURCE:

- (A) LIBRARY: GenBank
- (B) CLONE: 1314346

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

```

Met Ala Met Ser Ser Gly Gly Ser Gly Gly Gly Val Pro Glu Gln Glu
 1              5              10              15
Asp Ser Val Leu Phe Arg Arg Gly Thr Gly Gln Ser Asp Asp Ser Asp
      20              25              30
Ile Trp Asp Asp Thr Ala Leu Ile Lys Ala Tyr Asp Lys Ala Val Ala
      35              40              45
Ser Phe Lys His Ala Leu Lys Asn Gly Asp Ile Cys Glu Thr Ser Gly
      50              55              60
Lys Pro Lys Thr Thr Pro Lys Arg Lys Pro Ala Lys Lys Asn Lys Ser
      65              70              75              80
Gln Lys Lys Asn Thr Ala Ala Ser Leu Gln Gln Trp Lys Val Gly Asp
      85              90              95
Lys Cys Ser Ala Ile Trp Ser Glu Asp Gly Cys Ile Tyr Pro Ala Thr
      100              105              110
Ile Ala Ser Ile Asp Phe Lys Arg Glu Thr Cys Val Val Tyr Thr
      115              120              125
Gly Tyr Gly Asn Arg Glu Glu Gln Asn Leu Ser Asp Leu Leu Ser Pro
      130              135              140
Ile Cys Glu Val Ala Asn Asn Ile Glu Gln Asn Ala Gln Glu Asn Glu
      145              150              155              160
Asn Glu Ser Gln Val Ser Thr Asp Glu Ser Glu Asn Ser Arg Ser Pro
      165              170              175
Gly Asn Lys Ser Asp Asn Ile Lys Pro Lys Ser Ala Pro Trp Asn Ser

```

				180					185				190		
Phe	Leu	Pro	Pro	Pro	Pro	Pro	Met	Pro	Gly	Pro	Arg	Leu	Gly	Pro	Gly
		195					200					205			
Lys	Pro	Gly	Leu	Lys	Phe	Asn	Gly	Pro	Pro	Pro	Pro	Pro	Pro	Pro	Pro
		210				215					220				
Pro	Pro	His	Leu	Leu	Ser	Cys	Trp	Leu	Pro	Pro	Phe	Pro	Ser	Gly	Pro
225					230						235				240
Pro	Ile	Ile	Pro	Pro	Pro	Pro	Pro	Ile	Cys	Pro	Asp	Ser	Leu	Asp	Asp
					245					250				255	
Ala	Asp	Ala	Leu	Gly	Ser	Met	Leu	Ile	Ser	Trp	Tyr	Met	Ser	Gly	Tyr
			260					265					270		
His	Thr	Gly	Tyr	Tyr	Met	Gly	Phe	Arg	Gln	Asn	Gln	Lys	Glu	Gly	Arg
		275					280					285			
Cys	Ser	His	Ser	Leu	Asn										
	290														